LBNL-2001209

BERKELEY LAB

Lawrence Berkeley National Laboratory

The effect of linear regression modeling approaches on determining facility wide energy savings

Yuting Chen, Ph.D. and Peter Therkelsen, Ph.D.

Energy Technologies Area

April, 2019

This page has been intentionally left blank.

DISCLAIMER

ACKNOWLEDGEMENTS

This work was supported by the Assistant Secretary for Energy Efficiency and Renewable Energy, Advanced Manufacturing Office, of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

This page has been intentionally left blank.

Table of Contents

1. Abstract 7
1. Introduction 8
2. Background 9
3. Methodology 16
4. Results 16
4.1. Statistically Identifying Valid Relevant Variables 16
4.2. Forecast Model 17
4.2.1. Modeling Construction 17
4.2.2. Modeling Prediction 22
4.3. Backcast Model 23
4.3.1. Modeling Construction 23
4.3.2. Model Prediction 25
4.4. Standard Condition Approach 26
4.4.1. Model Prediction 27
4.5. Chaining Model 28
4.5.1. Model Construction 28
4.5.2. Model Prediction 29
5. Discussion 30
5.6. Model Comparison and Model Selection 30
5.6.1. Comparison of Energy Performance Improvement Percentage 30
5.7. New Energy Savings Indicator 34
5.8. Energy Performance Indicator with Confidence Interval 34
6. Conclusion 35
7. Bibliography 36
8. Nomenclature 37

This page has been intentionally left blank.

Abstract

Holistic energy management system business practices, such as the framework detailed in *ISO 50001 – Energy management system, requirements and guidance for use*, are centrally based upon the concept of energy performance improvement. For the purposes of ISO 50001, energy performance improvement can be determined for boundaries[1] and following a process that is best suited for the implementing organization. Many organization and government ISO 50001 based programs, including the United States Department of Energy Superior Energy Performance (SEP) program, find value in demonstrating organizational energy performance improvement as the difference in energy consumption between two time periods within physically defined facility-boundaries.

To make the difference in energy consumption for the two time periods meaningful, the amount of energy consumed must be adjusted to account for relevant variables. Relevant variables, such as metrics of production and weather conditions, affect directly the amount of energy consumed but are independent of the facility's energy performance. Therefore, it is crucial to adjust the observed energy consumption by the relevant variables identified by the facility, so that the energy savings resulted from the energy performance improvement actions can be isolated and determined, which is the purpose of this program.

To tackle this adjustment issue, the SEP measurement and verification (M&V) protocol specifies four energy consumption adjustment modeling methods for use; forecast, backcast, standard conditions, and chaining. Application of a single set of energy consumption and relevant variable data from a manufacturing facility to the four different energy consumption adjustment modeling methods produces four different energy savings values. Variation in the energy savings values is the result of inevitable changes in operation and conditions between the baseline and reporting periods, which affects the evaluation results significantly. The lack of agreement in the calculated energy savings values, while all meeting the requirements of the SEP M&V Protocol, indicates that additional context and analysis is required to understand which modeling method, and subsequent result, best represents the actual energy performance improvement of an organization.

This report describes how each adjustment model method can be implemented and provides guidelines of how to choose an appropriate adjustment method. A variety of statistical tests were made use to reveal which of the four methods best reflects the energy performance improvement of a given organization. In the study case of this paper, all of the four adjustment methods were applied. The resulting four savings estimates ranging from -1091.4 to 142,248.0 MMBtu, and the four SEP Energy Performance Indicator (SEnPI) estimates ranging from 0.93 to 1.00 lead to drastically different energy performance improvement conclusions. Discussion

[1] Facility boundaries are required to be established for which energy performance improvement value can be evaluated. Facility boundaries are considered three-dimensional and unchanged during the evaluation, so that the energy accounting shall include energy that enter the facility boundaries from the sky and ground if consumed at the facility.

focused on why an organization should pick one of the four modeling methods over the others is carried out. An average energy saving estimator was proposed in order to provide a more appropriate estimate for the ESP. The interpretation of the 95% confidence intervals associated to ESP may provide more context to what calculated energy savings values mean and how they should be interpreted.

1. Introduction

ISO 50001 based energy management systems require the demonstration of improvements in energy performance. Organizations implementing the international standard are empowered to select the energy performance indicators (EnPI) and approach to determining energy performance improvement that are suitable to their energy consumption mode and valuable for their energy consumption performance management.. As a result, an organization may employ one or more EnPIs based on different hierarchy levels for its own energy management needs or for certification to ISO 50001. These EnPIs can encompass any range of physical boundaries and scope, from being focused on individual equipment and systems to the organizational boundaries and all energy use within.

Many national energy efficiency programs based upon ISO 50001 specify a single EnPI that must be used when reporting energy performance improvement. The pre-established EnPI is typically used to determine normalized energy savings for an entire facility, rather than for individual systems and equipment within the facility boundary. The use of a single pre-defined facility-wide normalized energy savings EnPI enables government programs to compare and aggregate energy savings resulting from ISO 50001 implementation, a useful exercise to understand the societal and economic impact of a national program.

Many national ISO 50001 based programs specify the EnPI and process by which facility-wide energy savings are determined in a Measurement and Verification protocol. The United States Department of Energy (DOE) operates two nationwide ISO 50001 based energy efficiency programs, “50001 Ready,” and “Superior Energy Performance”. Both programs recognize organization for the implementation and continued use of an ISO 50001 energy management system, with SEP requiring third party certification to the international standard and 50001 Ready requiring self-attestation that the major components of an energy management system based upon ISO 50001 have been met. Energy performance improvement for both programs must be demonstrated at the facility boundary level. SEP requires use of the “Superior Energy Performance Measurement & Verification Protocol,” (SEP M&V Protocol) which details the verifiable process by which facility-wide, normalized energy savings for an organization are determined utilizing a single EnPI.

The SEP M&V Protocol allows for four methods by which organizations can normalize energy consumption to establish comparable time periods for use in calculating energy savings. These methods detail how to create energy consumption adjustment models normalize energy consumption for variations in values of relevant variables that affect the amount of energy consumed by an organization. Typical relevant variables used to develop energy consumption adjustment models for industrial facilities include operating conditions including production level, operating hours and weather conditions.

The four allowable energy consumption adjustment model methods deduce different energy savings values as they each specify the use of different time periods on which to build the adjustment model. Since the energy consumption adjustment model is a function of the relevant variables that affect energy consumption and their relationship to energy consumption values, the use of different time periods which contain different variation levels of these data points creates variations in the resulting model. These magnitude and spread of these values across a time period can be fairly different between the baseline period and the final evaluation period when consumption data are collected for evaluation use.

The allowance of multiple methods creates flexibility, enabling organizations to more readily participate in the program. However, application of a common data set to the four allowable methods usually result in different energy savings values. This paradox of one set of data resulting in multiple allowable results can lead those seeking to report the most accurate energy savings value to ask which allowable method should be used. The SEP M&V Protocol does not provide guidance or requirements pertaining to how to know which of the four normalization methods to select when determining energy savings. This leaves the SEP program prone to facilities selecting a modeling method that leads to the most favorable result, even if that result is not the most accurate representation of the savings achieved.

This report details and contextualizes the four energy consumption adjustment model methods allowed by the SEP M&V Protocol, demonstrates the variability of energy savings estimates that can occur when a single set of real industrial customer data is applied to each of the four methods, discusses why the methods generate different results, proposes that statistical tests can be employed in future works to guide model construction, and provides recommend to facilities which of the four methods should be used to determine energy savings based on relevant variable ranges and observation conditions.

2. Background

The determination of energy consumption is conducted with various purposes, such as to comply with legal standards, to meet sustainability requirements or to quantify potential energy savings. *ISO 50001 – Energy management system standard – Requirement with guidance for use*, is the internationally developed framework for an energy management system, a set of business practices used with the goal of improving energy performance improvement. The method for demonstrating energy performance improvement is left up to the organization implementing ISO 50001. The flexibility of ISO 50001 allows organizations to select energy performance indicators (EnPIs) that span different physical boundaries and scopes. Many organizations use some form of energy savings EnPI as their selected metric to demonstrate energy performance improvement and use either a standardized or custom approach to determining the energy savings value.

Government run ISO 50001-based programs have an interest in being able to compare and aggregate energy savings from participating facilities. To enable this, measurement and verification (M&V) protocols with EnPIs that represent the facility-wide energy savings are developed. The United States Department of Energy developed and maintains the Superior Energy Performance (SEP) M&V Protocol which requires use of such a facility-wide EnPI. The SEP M&V Protocol's EnPI is determined on a facility-wide (sometimes known as a top-down) basis that includes four methods to normalize energy consumption data to make energy data for a baseline and reporting period comparable. The four normalization methods, forecast, backcast, standard conditions, and chaining, all adjust either the baseline, reporting, achievement, or both baseline and reporting periods energy data with regression models built around relevant variables, independent factors that affect the amount of energy consumed but are not a result of energy performance improvement actions, such as production level and the weather conditions.

The EnPI required to quantify the energy performance improvement for the SEP program is the SEP Energy Performance Indicator (SEnPI). SEnPI is calculated as the facility-wide reporting period total primary energy consumption over the baseline period total primary energy consumption where the reporting ($\mathrm{ECP}(\Sigma)_\mathrm{r}$), baseline ($\mathrm{ECP}(\Sigma)_\mathrm{b}$), or both periods energy data are adjusted.

Equation 1 $$\mathbf{SEnPI} = \frac{\mathbf{ECP(\Sigma)_r}}{\mathbf{ECP(\Sigma)_b}}$$

An SEnPI value less than 1 suggests that the facility's energy performance is improved. The calculated SEnPI value is used to calculate energy performance improvement as a percentage value as = (1-SEnPI) x 100.

However, oftentimes $\mathrm{ECP}(\Sigma)_\mathrm{r}$ and $\mathrm{ECP}(\Sigma)_\mathrm{b}$ are measured under different circumstance that they cannot be compared directly. This is a result in differences in relevant variables such as production and weather between the two periods. Relevant variables usually affect energy consumption in a significant way and if not reflected in the energy performance improvement calculation would lead to misinterpretation regarding the impact of energy performance improvement actions taken with the objective of improving the facility's energy consumption performance. Therefore, an adjustment step is necessary to provide comparable energy consumption data to represent the facility's consumption level under baseline operation basis and under reporting period operation basis.

A regression adjustment model should be established that describes energy consumption as a function of relevant variables for each energy type included in the energy accounting. For industrial facilities that search for SEP certification, the energy consumption is usually considered in a linear relationship with the explanatory variable(s).

In most of the cases, the range of values for a given relevant variable varies from the baseline period to the reporting period. Data periods with too little or too great of resulting data variability can result in adjustment models that do not well represent the energy related performance across all time periods, and in some instances may result in the inability to create an allowable model for use to calculate energy savings that are allowable by some set of statistical tests outlined in an M&V Protocol. In the case of the SEP M&V Protocol, four different methods of model development are allowed to increase the likelihood that an allowable model can be constructed. Each method allowed by the SEP M&V Protocol dictates which data period (baseline, reporting, or an intermediate) should be used to create the energy adjustment model.

Being able to determine which method and subsequent data period as the range of comparison for the energy performance determination is crucial to developing a model that best reflects the energy performance of an organization. In the SEP M&V Protocol, four methods for selecting which data period on which to construct an energy consumption adjustment model. Table 1 displays a summary of the SEP M&V Protocol allowable energy adjustment model methods and their corresponding data periods on which the model would be developed.

Table 1 Summary of Adjustment Model Methods

		Primary Methods			Chaining Method	
		Forecast	Backcast	Standard Conditions	Baseline to Intermediate Period	Intermediate to Reporting Period
Energy Consumption Quantity	Reporting Period $\mathrm{ECP}(\Sigma)_{r}$	Observed (actual) $\mathrm{ECP}(\Sigma)^{o}_{r}$	Adjusted to baseline conditions $\mathrm{ECP}(\Sigma)^{a}_{r\|b}$	Adjusted to standard conditions $\mathrm{ECP}(\Sigma)^{a}_{r\|s}$	NA	Observed (actual) $\mathrm{ECP}(\Sigma)^{o}_{r}$
	Intermediate Period ECP(Σ)i	NA	NA	NA	Adjusted to baseline conditions $\mathrm{ECP}(\Sigma)^{a}_{i\|b}$	Adjusted to reporting period conditions $\mathrm{ECP}(\Sigma)^{a}_{i\|r}$
	Baseline Period $\mathrm{ECP}(\Sigma)_{b}$	Adjusted to reporting period conditions $\mathrm{ECP}(\Sigma)^{a}_{b\|r}$	Observed (actual) $\mathrm{ECP}(\Sigma)^{o}_{b}$	Adjusted to standard conditions $\mathrm{ECP}(\Sigma)^{a}_{b\|s}$	Observed (actual) $\mathrm{ECP}(\Sigma)^{o}_{b}$	NA
SEnPI Equation		$\frac{\mathrm{ECP}(\Sigma)^{o}_{r}}{\mathrm{ECP}(\Sigma)^{a}_{b\|r}}$	$\frac{\mathrm{ECP}(\Sigma)^{a}_{r\|b}}{\mathrm{ECP}(\Sigma)^{o}_{b}}$	$\frac{\mathrm{ECP}(\Sigma)^{a}_{r\|s}}{\mathrm{ECP}(\Sigma)^{a}_{b\|s}}$	$\frac{\mathrm{ECP}(\Sigma)^{a}_{i\|b}}{\mathrm{ECP}(\Sigma)^{o}_{b}} \times \frac{\mathrm{ECP}(\Sigma)^{o}_{r}}{\mathrm{ECP}(\Sigma)^{a}_{i\|r}}$	

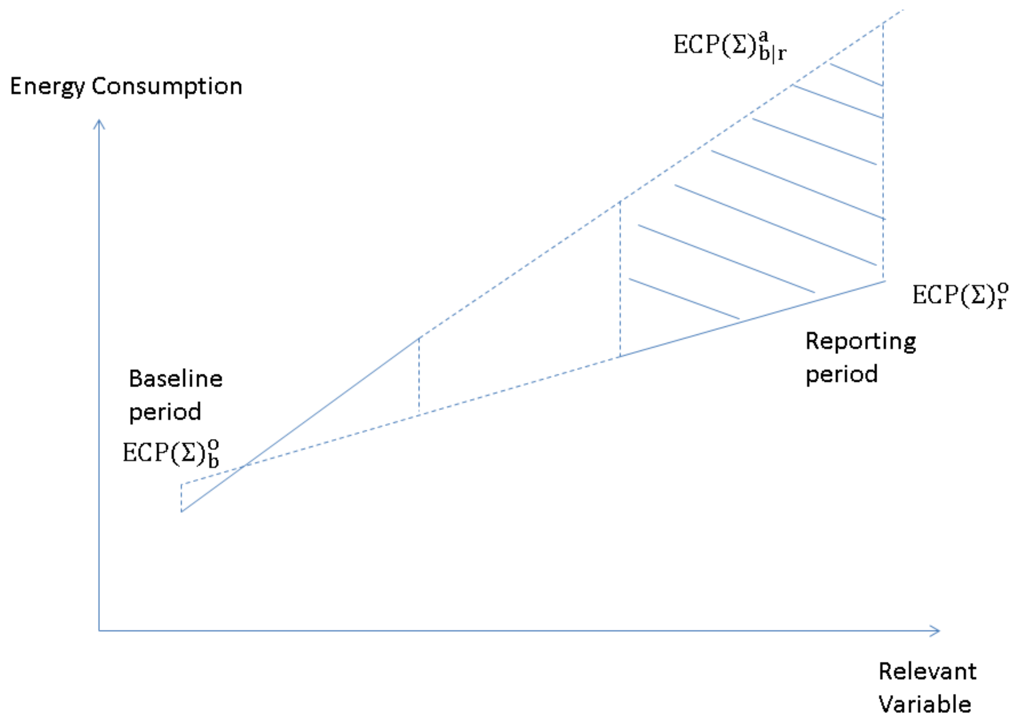

Figure 1 Forecast Model Framework

The forecast model method uses baseline period energy consumption as training data to describe the facility's energy consumption behavior during the baseline period. The obtained regression model is then applied to the reporting period relevant variable values to estimate the energy consumption that would have been expected at reporting period if the baseline operating systems and practices were carried into the reporting period. Energy performance improvement is calculated by comparing the model estimated energy consumption with the observed energy consumption at the reporting period.

Figure 1 illustrates a case when a relevant variable with important variations between the baseline and the reporting periods is chosen, how the energy performance improvement is calculated with forecast model. The shaded area represents the estimated energy savings.

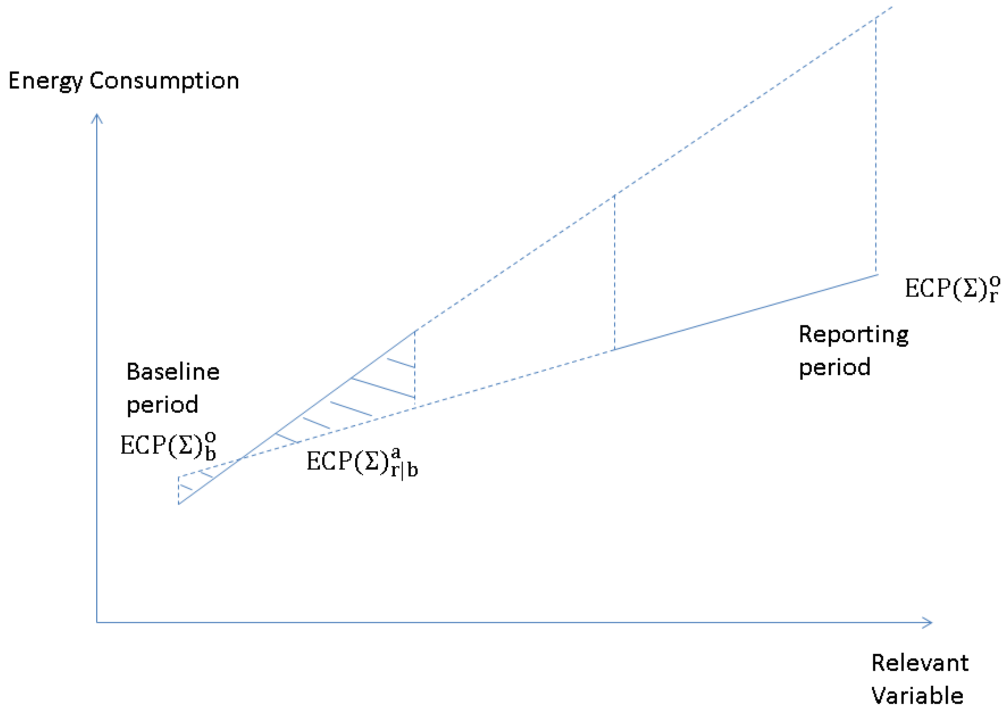

Figure 2 Backcast Model Framework

The backcast model method uses reporting period energy consumption data as training data to describe the facility's consumption behavior during the reporting period. The obtained regression model is then applied to the baseline relevant variable values to estimate the energy consumption that would have been expected at baseline period if the reporting period operating systems and practices were in place. By comparing the model estimated energy consumption with the observed energy consumption at the baseline period, the consumption performance improvement can be assessed. Figure 2 demonstrates how the backcast model estimates the energy performance and the energy savings based on the same data used for forecast model (shown in Figure 1). The shaded area represents the estimated energy savings.

The standard conditions model method is established with the basis of two regression models; a forecast model fitted with baseline period data, and a backcast model obtained relying on reporting period data. Standard conditions (relevant variable values) are applied to both regression models resulting in two sets of energy consumption estimtes that describes what level of energy consumption would have been expected for these standard conditions, one value describing if baseline period operating system and practices were in place and the other describing if reporting period operating system and practices were in place. A case of standard conditions model is shown in Figure 3, in which the same set of energy consumption data used in the forecast and backcast models (shown in Figure 1 and Figure 2 respectively). The estimated energy savings are consist of the two shaded areas which are the difference of energy consumption projections obtained with the forecast and the backcast models.

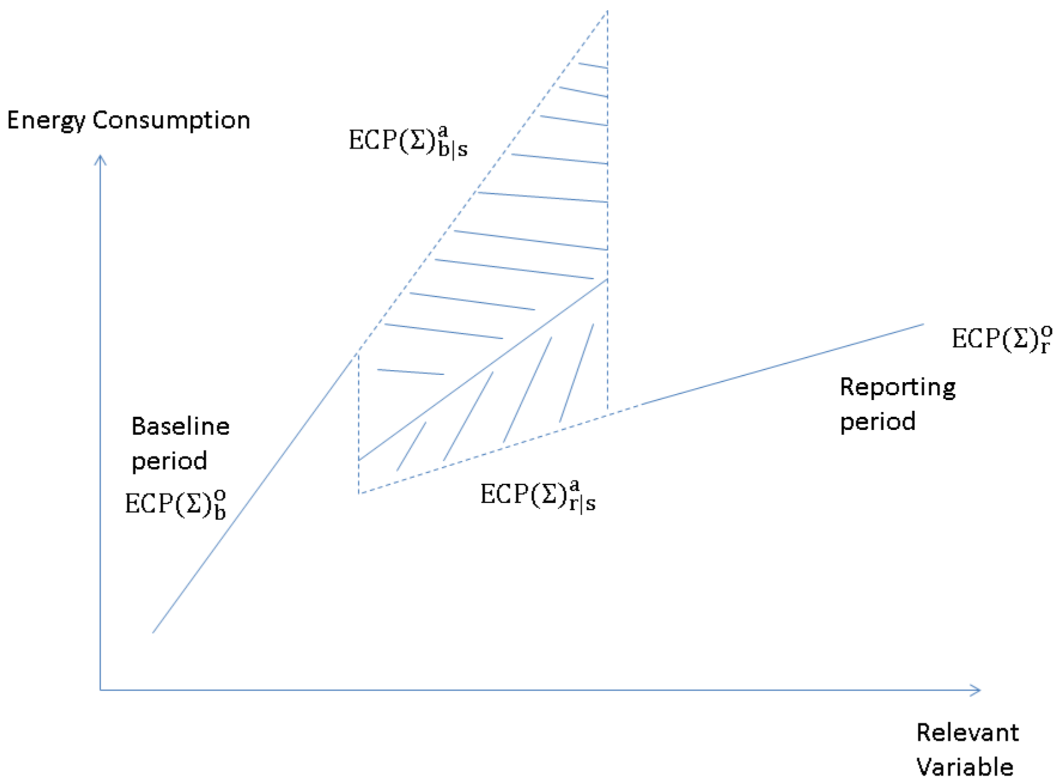

Figure 3 Standard Conditions Approach Framework

The chaining model method is a composite of the forecast and backcast methods. The chaining method may be used if there is an intermediate period between the baseline period and the reporting period for which a model can be developed. In the case that the data were collected according to the SEP M&V protocol, this intermediate period must be at least one year long. The energy consumption model developed for the intermediate period is used to backcast to the baseline period, and to forecast to the reporting period to obtain the two sets of energy consumption estimates. The differences between the estimated consumption and the observed consumption data are the estimates of the energy consumption savings.

Chaining is useful if one of the following is true:

1) the reporting period conditions (relevant variable range) for energy consumption measurements are very different from the conditions of the baseline period model,
2) the intermediate period between the baseline period and the reporting period is longer than 36 months,
3) data availability or quality issues exist for either the baseline or reporting periods,
4) a forecast or backcast adjustment model cannot be developed for either the reporting or baseline periods,
5) the intermediate period relevant variable values are close to the future operating condition as compared to both the baseline and reporting periods.

Figure 4 illustrates a toy example of how chaining model can be used to evaluate energy performance when the relevant variable chosen has significant gaps between baseline period and reporting period. The estimation is performed on the same dataset used with forecast, backcast and standard conditions models. The sum of the two shaded areas located in the reporting period zone are the estimated energy savings.

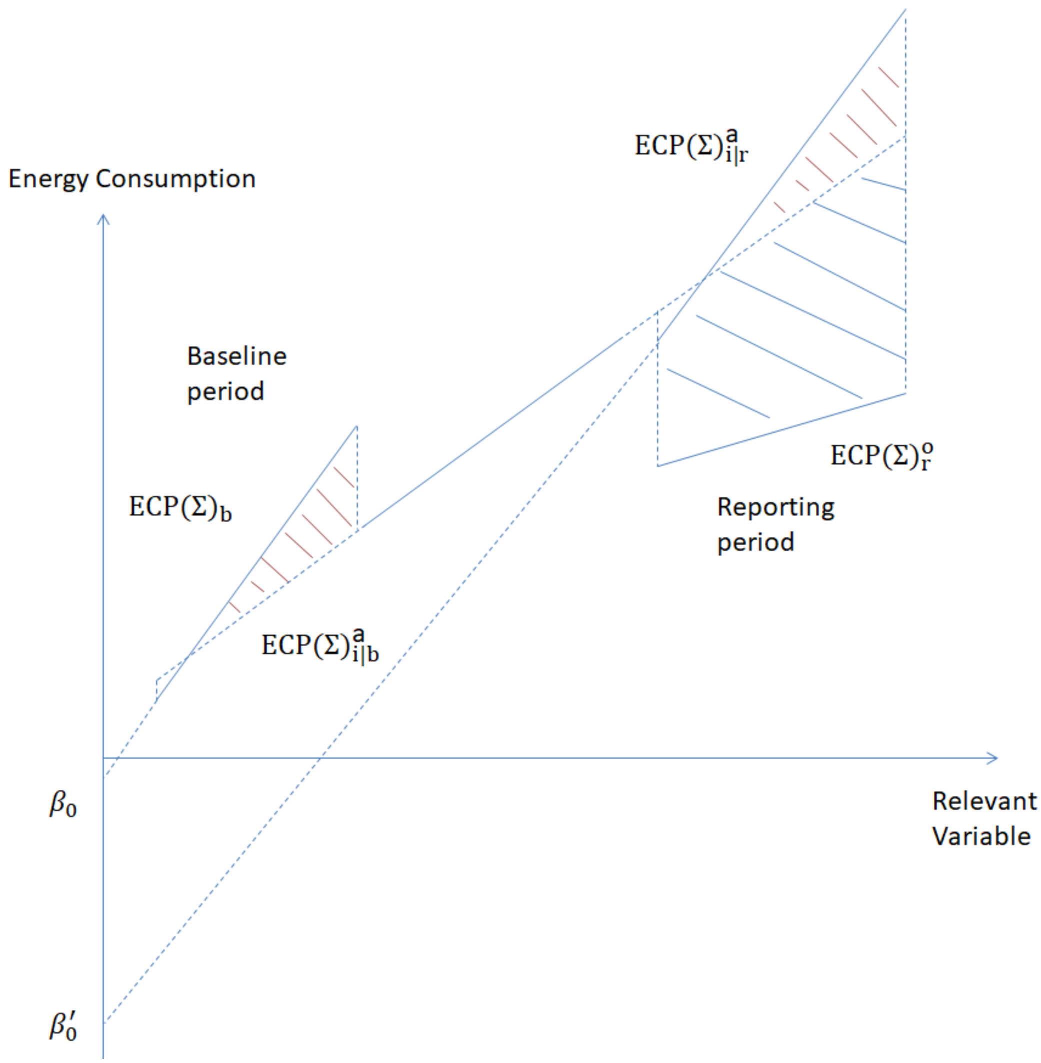

Figure 4 Chaining Model Framework

The allowance of the four modeling methods in the SEP M&V Protocol allows for flexibility and greater program participation. The M&V Protocol lacks guidance on determining which method is considered most appropriate and should be chosen for facilities with versatile background and energy consumption data obtained under various conditions, while such liberty may introduce significant difference in the estimation of energy savings which may lead to drastically different conclusion regarding energy performance improvement. This caveat leaves the SEP program prone to facilities selecting a method that provides most favorable result, even if that result is not the most accurate representation of the savings achieved.

In this paper, the four methods are applied to the energy consumption and relevant variable data provided by a sample manufacturing facility in the United States. Estimation results from each method are presented and compared, and approach selection criteria are suggested.

3. Methodology

In this paper, a common set of energy consumption and relevant variable data from a manufacturing facility in the United States is applied to all the four energy consumption adjustment models allowed by the SEP M&V Protocol. The results are analyzed with a number of statistical tests elaborated to determine if the assumptions based on which the regression models are built are satisfied, and to evaluate how an organization can understand which of the four methods renders the greatest meaning for its specific set of data.

The SEP M&V Protocol has gone through a number of revisions and the March 8, 2017 version is referenced in this study. There are no differences in the allowable four modeling methods between the March 2017 and any of the previously published versions of the SEP M&V Protocol.

The data used in this report comes from a manufacturing facility that was certified to the SEP program. The facility voluntarily supplied data pertaining to its facility energy consumption and relevant variable data. The facility consumes both natural gas and electricity (all in MMBtu). Outside of an initial review of the data only electricity energy consumption is used throughout the report, since the electricity consumption analysis exercise is already fairly representative of the crucial model selection and result interpretation issues that commonly manifest in the data analysis for majority of facilities participating the SEP program and require a high degree of attention. As a result, the natural gas consumption analysis is not included to avoid the paper being too lengthy.

Relevant variables include heating degree days (HDD) and a production metric which has been masked as “A”. All data were collected on a monthly basis across four years. The first year of data constitutes the baseline period with the next three years making up the achievement period. The final year of the achievement period is referred to as the reporting period. These time periods are specific to the SEP program.

A discussion of the four modeling methods is followed by the application of confidence intervals and the development of an alternative energy savings metric based upon the standard conditions method and taking into account the full range of relevant variable data available.

4. Results

4.1. Statistically Identifying Valid Relevant Variables

To build a reliable regression model, first we need to identify the reasonable independent variables that result in the outcome. Figure 5 provides plots of energy consumptions and related relevant variable data provided by the manufacturing facility spanning all data periods (baseline, achievement, and reporting).

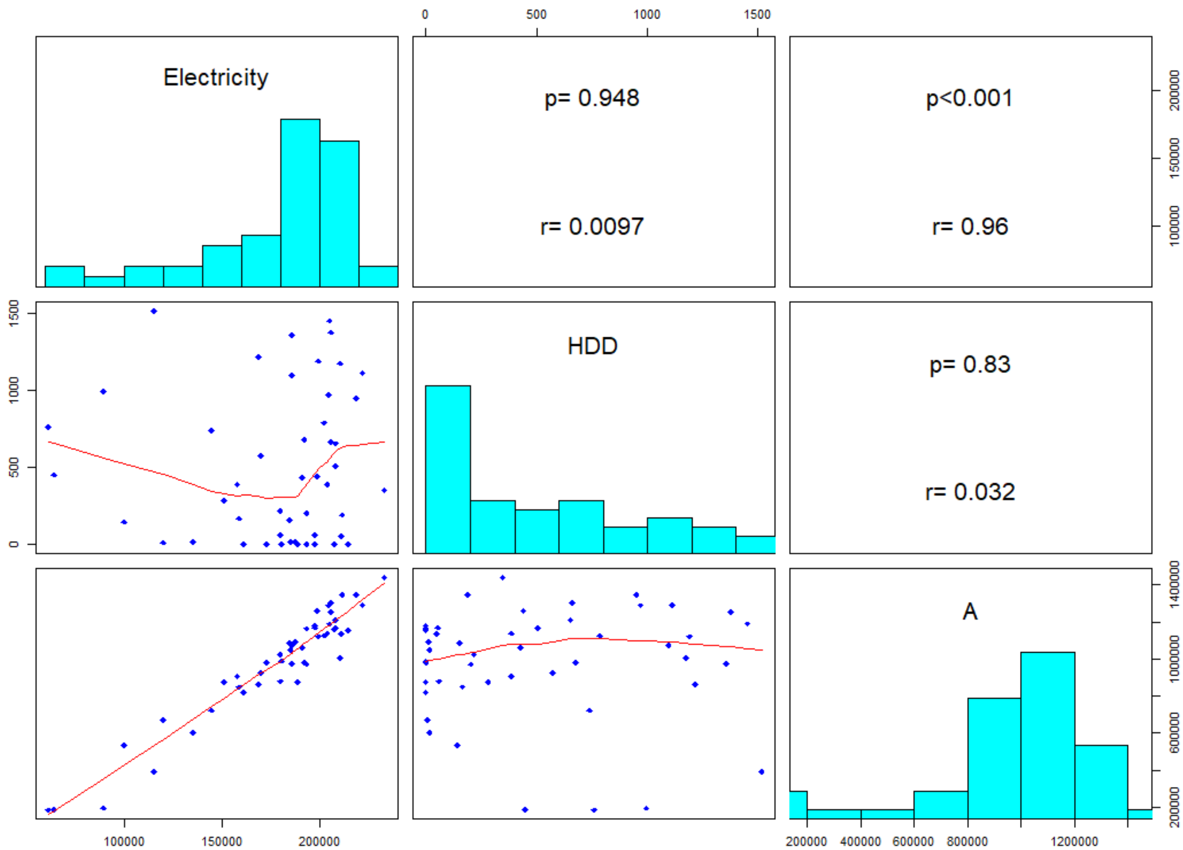

Figure 5 Correlation between Energy Consumption (in MMBtu) and Independent Variables

Evaluating for electricity, even a cursory visual analysis of Figure 5 reveals that relevant variable "A" is strongly correlated to the consumption of electricity. The Pearson's product-moment correlation, also known as correlation coefficient, which is often employed as a measure of the strength of linear association between two variables with 0 indicating no association and ±1 indicating the strongest positive or negative association, shows a value of 0.96, which mathematically confirms the assumption of a strong positive correlation between the dependent variable (Electricity) and the independent variable ("A"). The four energy consumption adjustment model methods will be used to construct electricity models with "A" as the relevant variable.

4.2. Forecast Model

4.2.1. Modeling Construction

A forecast model using baseline period relevant variable and energy consumption data is constructed.

Equation 2 $$\boldsymbol{Electricity_{baseline} = \beta_o^F + \beta^F_{"A"} X^{baseline}_{"A"} + \varepsilon_i^F}$$

$Electricity_{baseline}$	= electricity consumption observed during the baseline period,
β_o^F	= forecast model intercept,
$\beta_{"A"}^F$	= coefficient of the independent variable "A" in forecast model,
$X_{"A"}^{baseline}$	= baseline period observed value of variable "A",
$\varepsilon_i^F \sim N(0, \sigma_{baseline}^2), i.i.d.$:	= forecast model error term.

The fitted model can be described as follows:

$$Electricity_{baseline} = \mathbf{53567.33} + 0.13 \cdot X_{"A"}^{baseline} + \varepsilon_i^F,$$

with $R^2 = 0.95$. This means that the model explains 95% of the variability of the observations (data). A p-value of 7.12e-8 for F-statistic indicates that the linear relationship is statistically significant (7.12e-8 < 0.05), and the model fits the data significantly better than the mean[2].

To ensure the model developed is rigorous, the following statistical tests are conducted to verify whether:

- The regression function is nonlinear
- The error terms are not normally distributed
- The error terms have non-constant variance
- The error terms are not independent
- There are outliers

Table 2 Assumptions and Verification Tests for Linear Model

Assumption	Diagnostic Methods	Consequence of Violation
Linearity	Plot of observed versus predicted values	Errors in prediction, especially when extrapolating
Normality of the error distribution of the residuals	Normal quantile plot	Bias in the determination of model coefficients and in the calculation of confidence intervals for forecasts
Homoscedasticity	Plot of residuals versus predicted values	Hard to evaluate the standard errors of the coefficients which will result in bias in estimated confidence intervals
Collinearity and independence of the errors	Plots of residuals for all observations; Variance inflation factor; Durbin-Watson statistic	If errors tend to always have the same sign, the model systematically under-predicts or over-predicts under particular configuration.
No outliers	Plots of Cook's distance and Studentized residuals	Outliers may be due to random variation or indicate areas that the linear model is not a good fit.

[2] A constant model y= average of all observed values.

Cook's distance is a common influence measure tool to identify influential data points (see Figure 6), which can potentially be outliers. An influential data point is defined as if removed from the data would significantly change the fit. It can be either an outlier or a point having large leverage[3] which implies that the independent variable of a point is far from those of other observations.

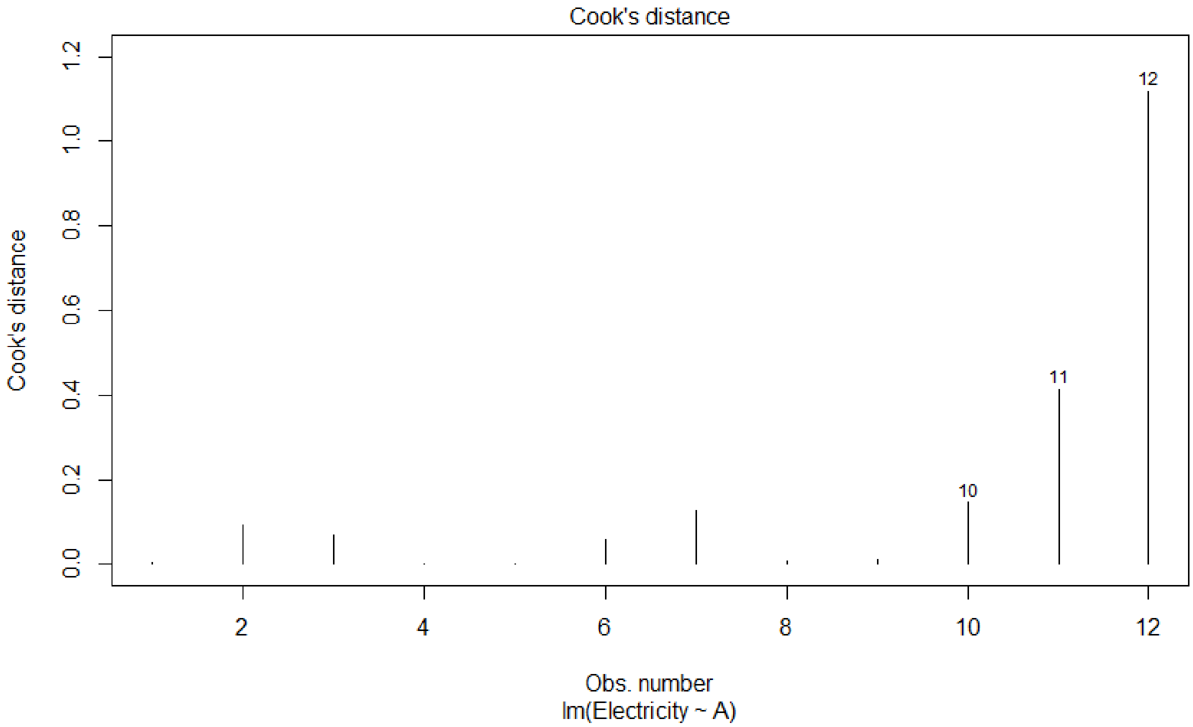

Figure 6 Cook's Distance of Observations of Reporting Period Based on the Forecast Model

The outlier test of Studentized residual indicated that no outliers were detected with a significant p-value of 0.05. The point with largest Studentized residual is point 12.

Figure 7 identifies that both point 11 and point 12 are observations made with extremely low variable "A". This may explain why they were considered as having important influence on the fitted line's slope. However, since their Studentized residuals are all in reasonable range (<3), [4] these points are not deemed as outliers and thus can be included in the model construction.

In the case that there are potential outliers identified by the test, further investigation is required to identify the nature of the outliers. If the outlier is due to a measurement error, which is related to the personnel involved or tool used, and can be confirmed that it is

[3] The average of leverage is (number of independent variables +1)/number of observations. A large leverage can be 2 or 3 times of this value.

[4]When taking a standard normal distribution, only 5% of standardized residuals are outside +-1.96. A standardized residual outside +-3, which corresponds to only 0.28% of all the standardized residuals is conventionally considered as an outlier, which means that the chance that such a point belongs to the group of other points is less than 0.28%.

mistaken, then the point can be removed from the analysis. However, the removal should be reported and justified in the documentation, and a sensitivity analysis should be performed (with and without the removal point).

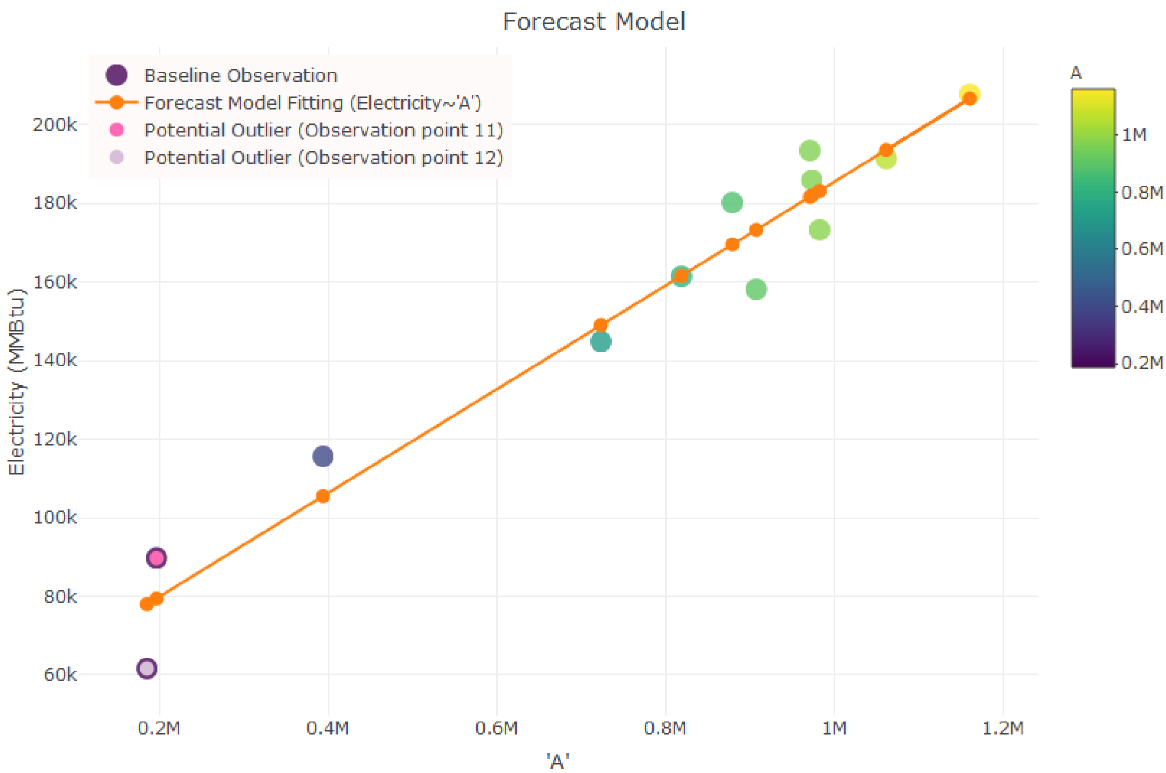

Figure 7 Forecast Model with Fitted Points and Baseline Period Observations

To check the normality of the error distribution, a histogram can be employed to illustrate the distribution of the residuals and to verify if the obtained residuals are symmetric and centered on zero, as shown in Figure 8. The estimated density is similar to a normal density distribution, therefore, the normality of the residuals can be assumed.

Q-Q plot is used to evaluate the normal assumption of the residuals. It compares the observed residual distribution to theoretical one by plotting their quantile against each other.

Figure 9 suggests that the error terms can generally be considered as normally distributed. Note that due to the limited size of the data points used to fit the model, the obtained results are contextual. If more data were collected, the conclusion could have been more reliable.

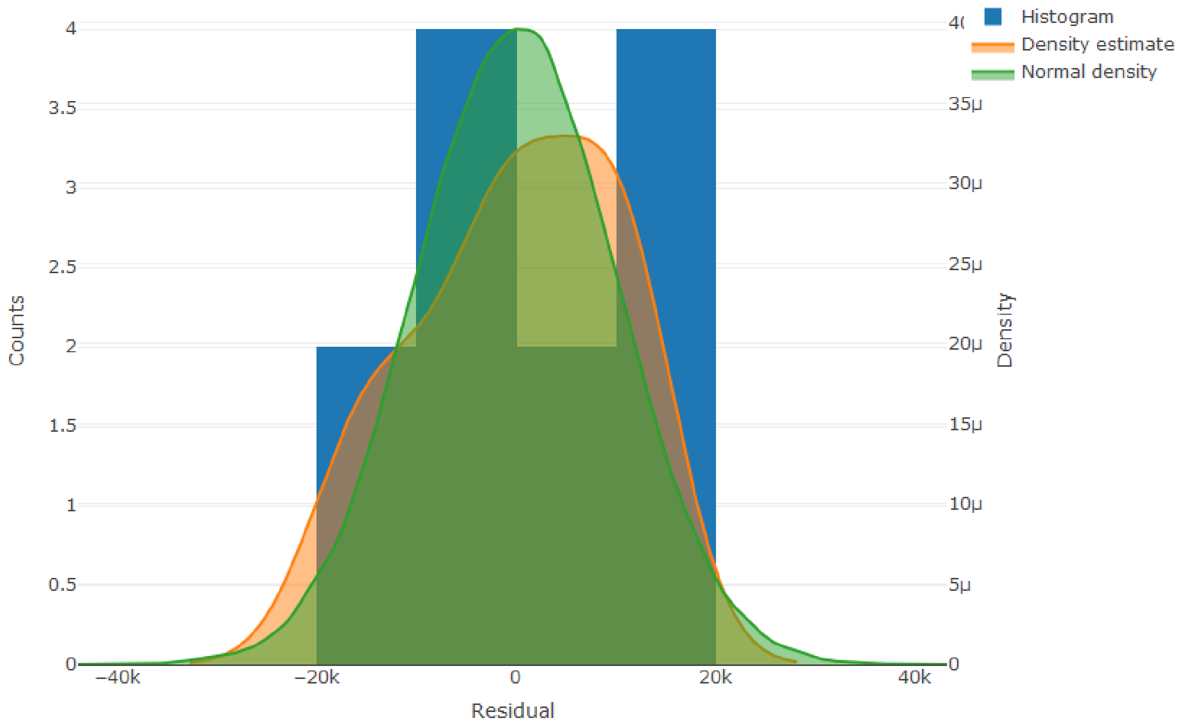

Figure 8 Histogram of the Residuals of the Forecast Model

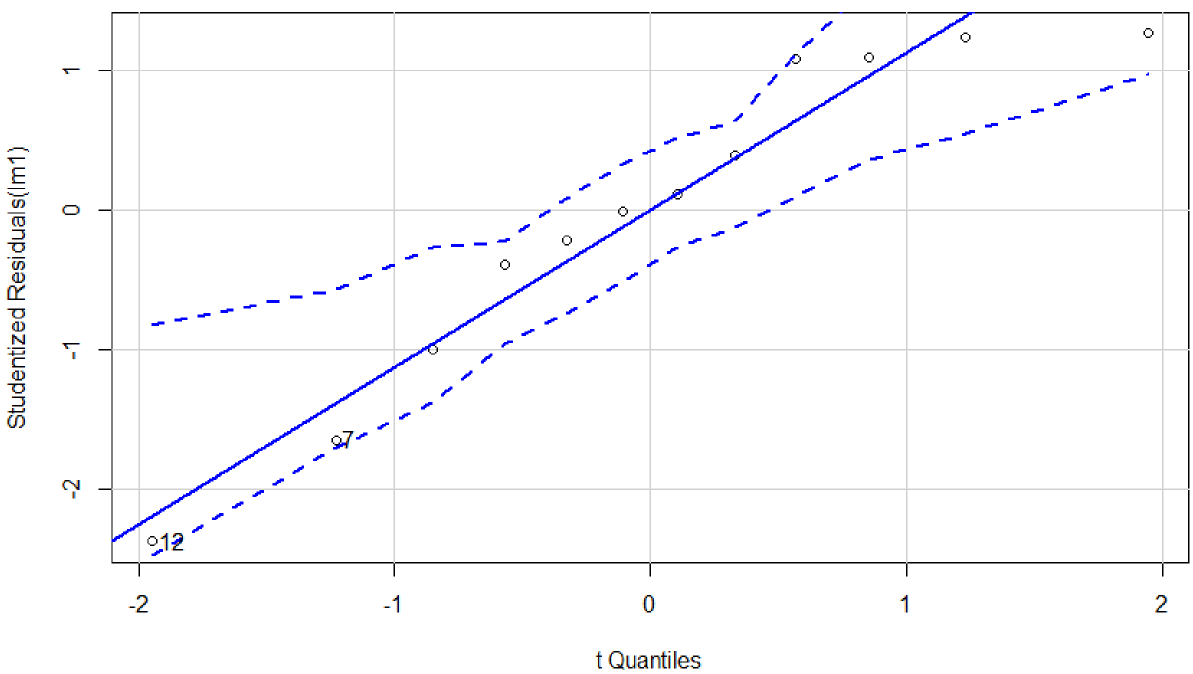

Figure 9 Quantiles Plot of Forecast Model

Another easier way to verify all the statistical constrains required by the linear model is to use the gvlma tool (package in R) to perform a global test of the linear model assumption, including assumptions on skewness and kurtosis of the residual distribution, link function (linear model statistically significant) and heteroscedasticity (the variance of the residuals is dependent on the independent variable – "A" in this case).

Table 3 Output of the gvlma tool (package in R) on the forecast model lm(Electricity ~ A)

```
ASSESSMENT OF THE LINEAR MODEL ASSUMPTIONS
USING THE GLOBAL TEST ON 4 DEGREES-OF-FREEDOM:
Level of Significance =  0.05

                     Value p-value                Decision
Global Stat        1.84877  0.7635 Assumptions acceptable.
Skewness           0.27453  0.6003 Assumptions acceptable.
Kurtosis           0.61915  0.4314 Assumptions acceptable.
Link Function      0.09512  0.7578 Assumptions acceptable.
Heteroscedasticity 0.85997  0.3537 Assumptions acceptable.
```

The results from the global test (gvlma) confirm that the linear model assumptions are all satisfied. We may use linear model to measure the reporting period electricity consumption based on the independent relevant variable ("A").

4.2.2. Modeling Prediction

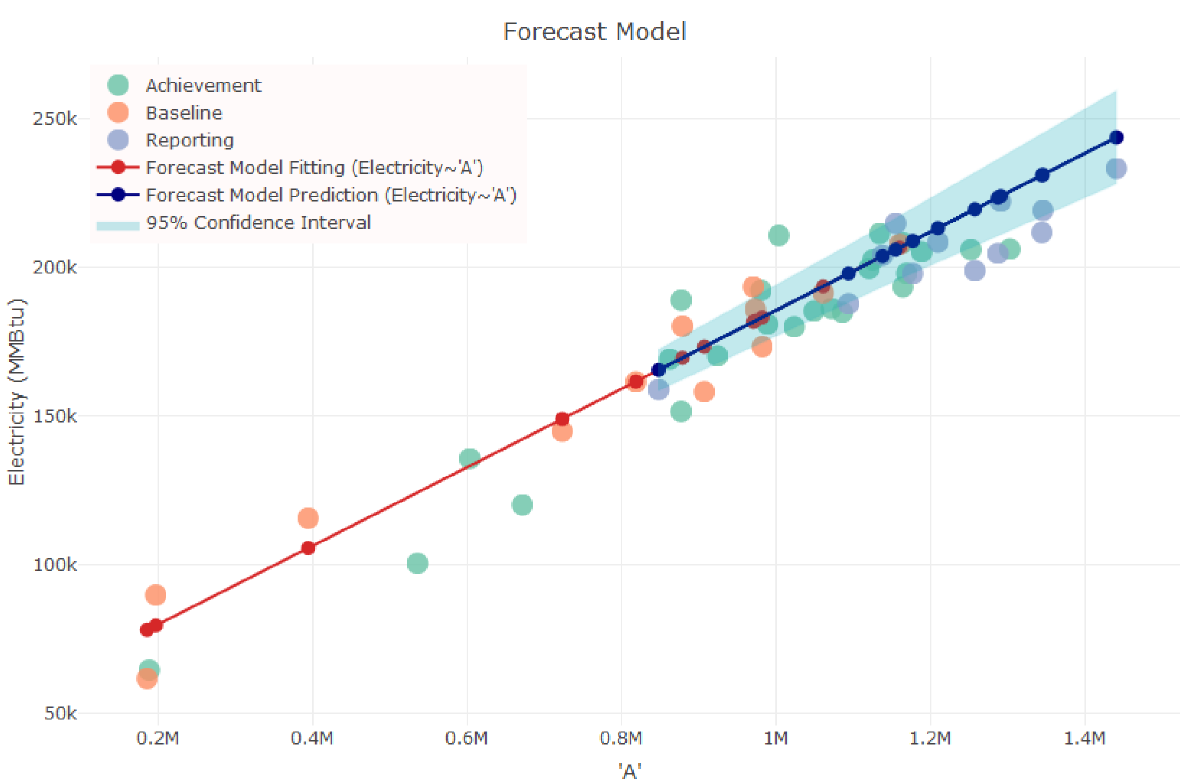

Figure 10 Reporting Period Energy Consumption Prediction of Forecast Model

Figure 10 illustrates the prediction results of the reporting period energy consumption from the established forecast model with 95% confidence interval.

Once the prediction is conducted for the reporting period, we could compare the predicted values with the observed reporting period electricity consumption. The predicted values in the reporting period are extrapolations, which imply that if the facility did not take actions to improve their energy performance, their energy consumption during the reporting period would have been the values predicted. Comparing the predicted values and the observations made in the reporting period may indicate whether the energy performance improvement of the facility is statistically significant. Given the fact that the sample size is relatively small, a Student's t-test is performed to examine whether the prediction mean and the observation mean are equal to each other (null hypothesis).

We obtain a p-value of 0.28, which is larger than the cutoff 0.05 (with 95% confidence level) indicating that there is weak evidence against the null hypothesis. Therefore, we cannot reject the null hypothesis in favor of the alternative hypothesis and could conclude that the predicted mean and the observed mean are not significantly different. The energy consumption performance of the facility during the baseline period and during the reporting period is not significantly different.

The above workflow demonstrates how to use a forecast model to evaluate and to quantify energy performance improvement for a given facility.

4.3. Backcast Model

4.3.1. Modeling Construction

A backcast model using the reporting period as the training data to fit our model and to predict the electricity consumption during the baseline period is constructed.

Equation 3 $$\boldsymbol{Electricity_{reporting} = \beta_o^B + \beta_{"A"}^B X_{"A"}^{reporting} + \varepsilon_i^B}$$

$Electricity_{reporting}$ = electricity consumption observed during the reporting period,
β_o^B = backcast model interception,
$\beta_{"A"}^B$ = coefficient of the independent variable "A" in backcast model,
$X_{"A"}^{reporting}$ = reporting period observed value of variable "A",
$\varepsilon_i^B \sim N(0, \sigma_{reporting}^2), i.i.d.$: = backcast model error term.

The fitted model is:

$$\boldsymbol{Electricity_{reporting} = 6917.97} + 0.11 \cdot \boldsymbol{X_{"A"}^{reporting} + \varepsilon_i^B,}$$

with $R^2 = 0.82$. This indicates that the model explains 82% of the variability of the observations (data). A p-value of 5.69e-5 for F-statistic implies that the linear relationship is statistically significant (5.69e-5 < 0.05), and the model fits the data significantly better than the mean.

Both outlier test and influence plot were carried out to identify potential outliers. To be qualified as an outlier, an observation point usually has important residual (unusual value of the dependent variable Y based on its independent variable X). Only in the case that an outlier has high leverage, which is measured by hat-value, will it influence strongly the regression's slope and intercept. In such case, the observation point must have an unusual X-value and an unusual Y-value given its X-value.

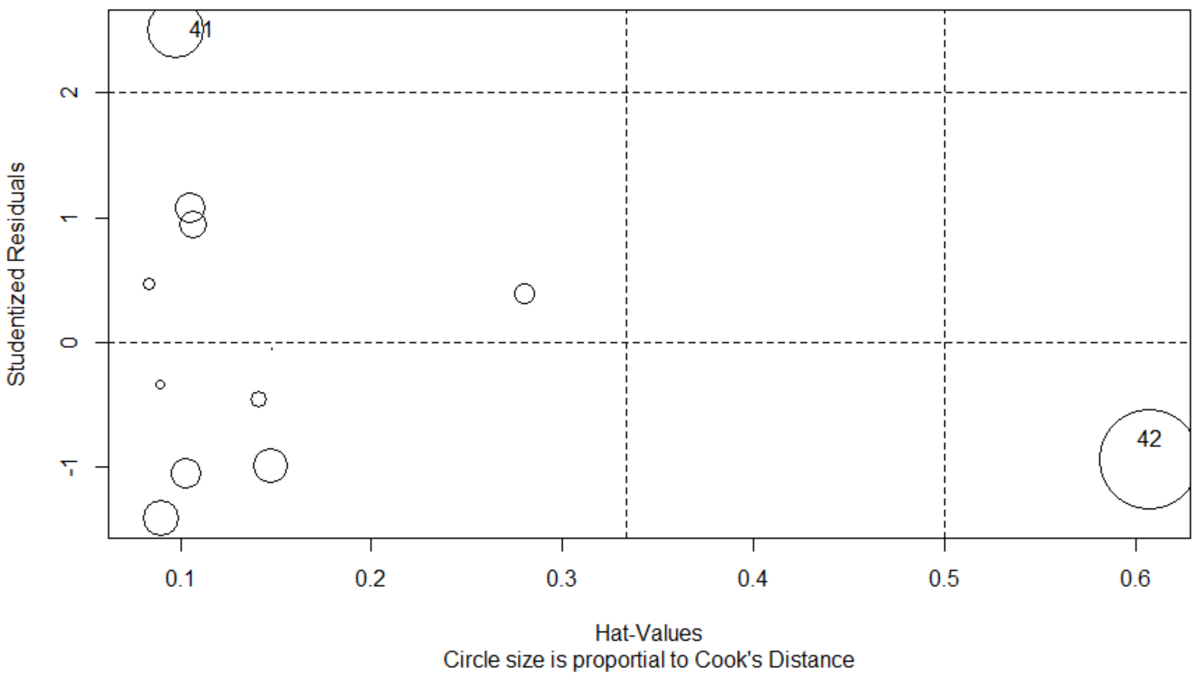

Figure 11 Influence Plot of Residuals of Backcast Model

Each point in Figure 11 represents an individual data observation. Although point 42 was evaluated as having larger influence than point 41 based on the influence plot in Figure 11, the Studentized residual test states that point 41 has the largest residual. No points were identified as an outlier since no Studentized residual is smaller than 3.

Figure 12 implies that point 42 was identified as an influential point because of its location. In any cases, measuring conditions are required to be verified in order to exclude any data points from the analysis.

The global assumption test confirms that all the linear model assumptions are all satisfied. We may use linear model to measure the reporting period electricity consumption based on the independent variable "A".

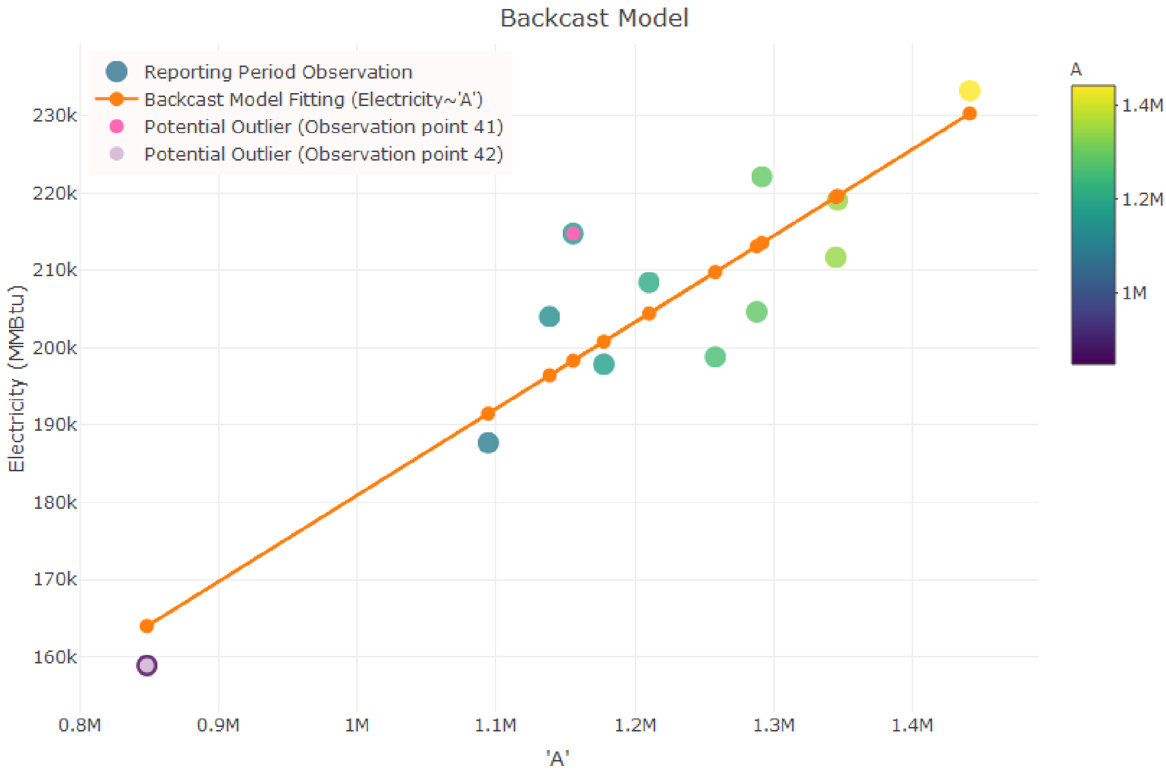

Figure 12 Backcast Model with Fitted Points and Reporting Period Observations

4.3.2. Model Prediction

Figure 13 illustrates the prediction results from the established backcast model with 95% confidence interval. Note that comparing to the forecast model in Figure 10, the backcast model is fitted with data with less variability. In other words, the relevant variable "A" covers a narrower range of values in the reporting period than in the baseline period. As a result, the backcast model's predictive capability is more limited than the forecast model, as shown in Figure 13, the estimated confidence interval covers a larger range in the backcast model.

In some cases, the consumption behaviors during the two evaluation periods may not be comparable due to the important difference between the relevant variable. In the event that the difference is larger than 3 times of the standard deviation, the predictive capacity would reach its limit and a Chaining model would be a better fit to perform the data normalization as well as the saving estimation. The application of Chaining model to the sample data is laid out in Section 4.5.

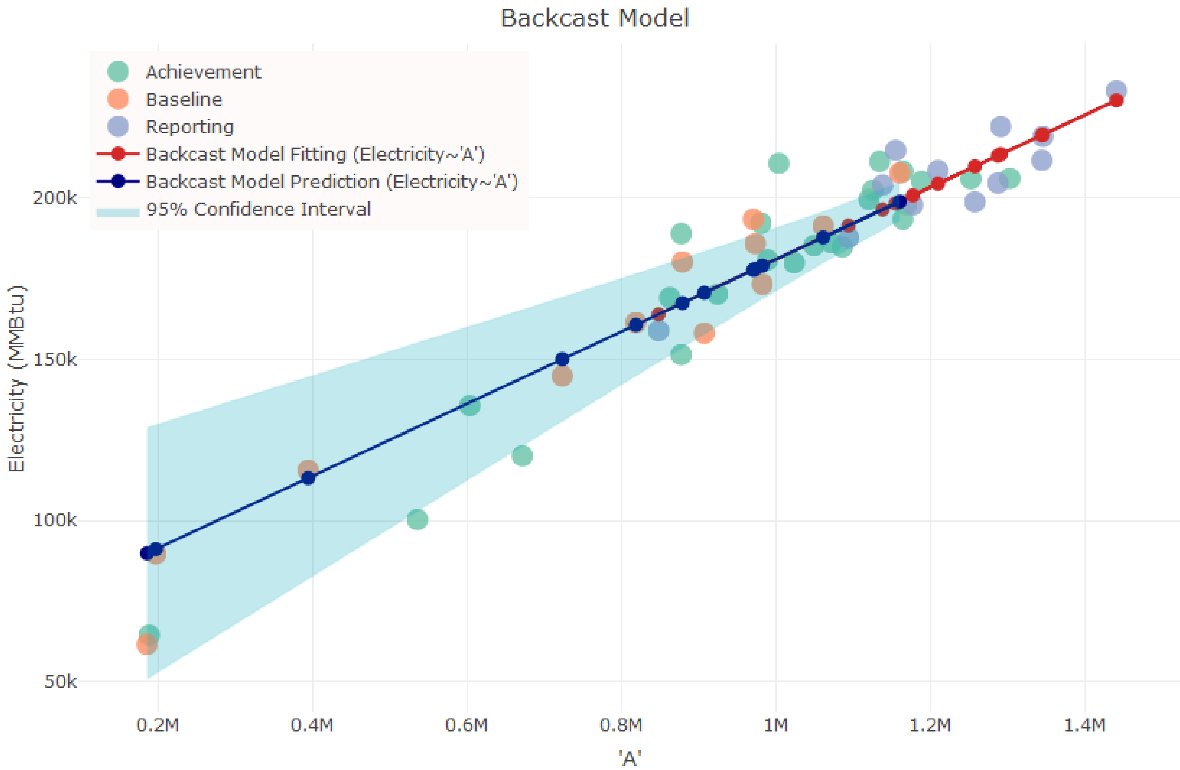

Figure 13 Baseline Period Energy Consumption Prediction of Backcast Model

As in the previous section, the projected consumption data can be compared to the observed baseline electricity consumption with a t-test to identify if the electricity consumption performances are statistically significantly different before and after the improvement actions.

The obtained p-value is 0.99 which suggests that we cannot reject the null hypothesis in favor of the alternative hypothesis and could conclude that the predicted mean and the observed mean are not significantly different. This conclusion matches the conclusion obtained from the forecast model analysis that the energy performances during the baseline period and the reporting period are not statistically significantly different (two consumption data sets may belong to the same population).

4.4. Standard Condition Approach

From the previous analysis, it is clear that the relevant variable ("A") range varies between the baseline and reporting periods. To enable the energy performance assessment based on a standardized level of values for the relevant variable "A", the forecast model (Equation 2) and backcast model (Equation 3) can combined to predict the energy consumption under the performances from baseline period and from reporting period respectively.

As shown in Figure 14, the achievement period, which is a period between the baseline period and the reporting period during which the energy performance improvement actions were taken, has data encompassing both low and high relevant variable "A" values and lasted for 2 years in this study case (24 data points). The wide spread of relevant variable makes this period well suited to represent the “standard conditions” of the facility. The previously constructed forecast and backcast models are used to predict the electricity consumption during the achievement period.

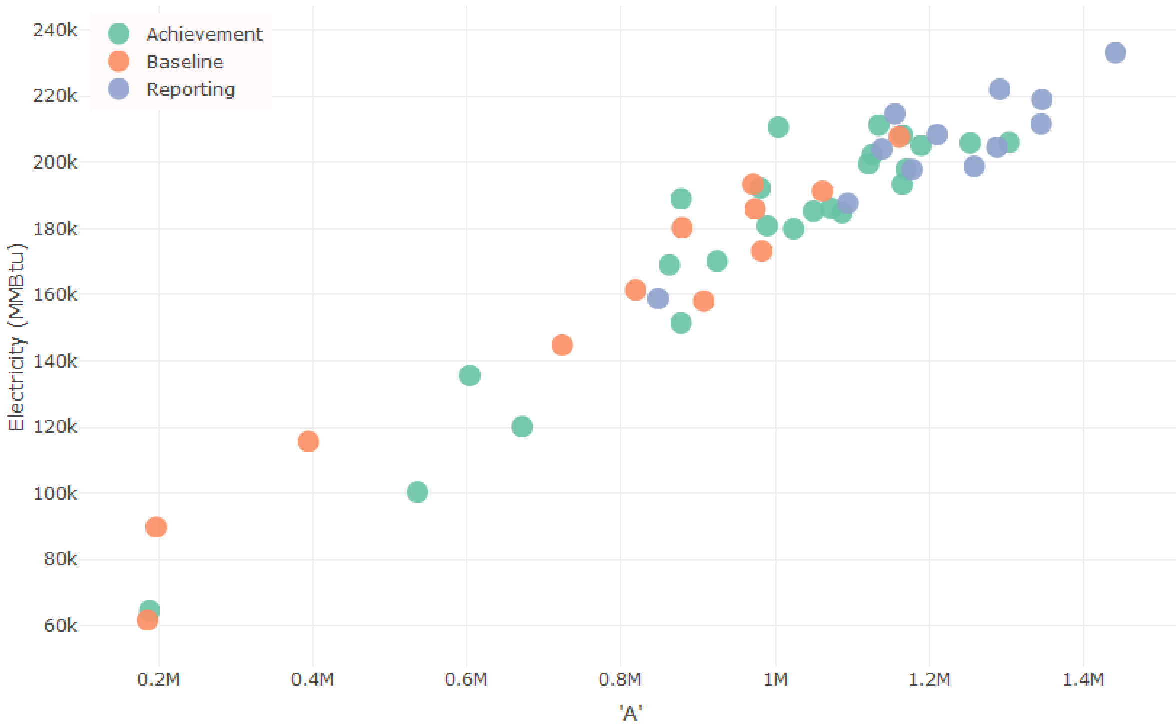

Figure 14 Electricity Consumption Observation from Baseline, Achievement and Reporting Periods

4.4.1. Model Prediction

Figure 15**Error! Reference source not found.** illustrates the prediction results from both the forecast and the backcast models based on achievement period production level with 95% confidence interval. Due to the lack of training data at low value range of relevant variable “A”, the backcast model showed important uncertainty for low value range of “A”, for it was established based on reporting period observations consisting only high value range of “A”.

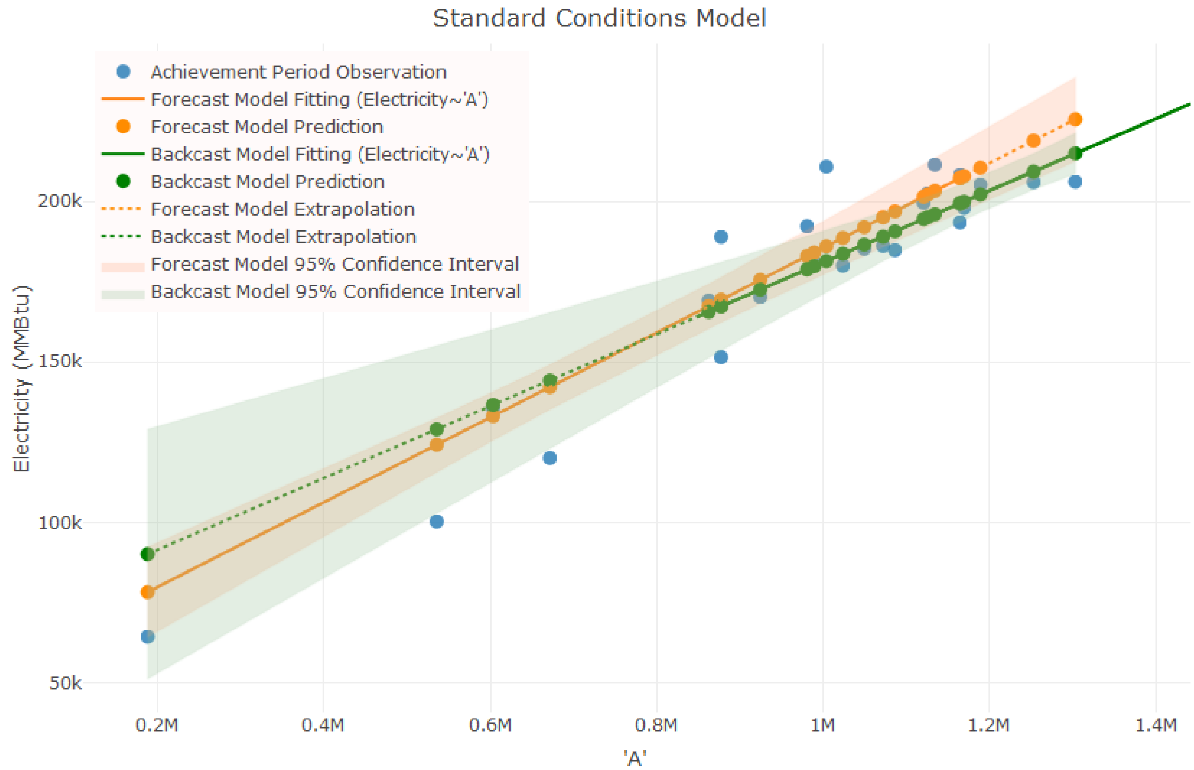

Figure 15 Achievement Period Energy Consumption Prediction of Standard Condition Approach with 95% Confidence Interval

A t-test is conducted to identify if the predicted consumption series during the achievement period by forecast model (orange points) and by backcast model (green points) are statistically significantly different in Figure 15. The obtained p-value 0.66 confirms the conclusion drawn from the previous t-tests that the energy consumption performances during the baseline and the reporting periods are not significantly different.

4.5. Chaining Model

4.5.1. Model Construction

Another way to evaluate the energy performance of baseline period and reporting period observations measured under different conditions via an intermediate period is the chaining model. A regression model is built based on observations made during an intermediate period, achievement period, which is between the baseline period and the reporting period. Lasting often longer than a year, the achievement period is usually capable of providing more training data points and covers larger range of relevant variables than the baseline or the reporting periods. This allows the chaining model to have broader applications.

Equation 4

$$\boldsymbol{Electricity_{achievement} = \beta_o^C + \beta_{"A"}^C X_{"A"}^{achievement} + \varepsilon_i^C}$$

$Electricity_{achievement}$ = electricity consumption observed during the achievement period,

β_o^C = chaining model interception,
$\beta_{"A"}^C$ = coefficient of the independent variable “A”,
$X_{"A"}^{achievement}$ = achievement period observed value of variable “A”,
$\varepsilon_i^C \sim N(0, \sigma_{achievement}^2), i.i.d.$: = chaining model error term.

The fitted model is:

$$\boldsymbol{Electricity_{achievement}} = 4211.16 + 0.14 \cdot \boldsymbol{X_{"A"}^{achievement}} + \boldsymbol{\varepsilon_i^C},$$

with $R^2 = 0.89$. This indicates that the model explains 82% of the variability of the observations (data). A p-value of 2.26e-12 for F-statistic implies that the linear relationship is statistically significant (2.26e-12 < 0.05), and the model fits the data significantly better than the mean.

Based on the outlier test conducted and the influence plot, no data points can be regarded as outliers. The global assumption test performed indicates that the linear model assumptions are all satisfied. We may use the linear model to measure the achievement period electricity consumption based on the independent variable “A”.

4.5.2. Model Prediction

Figure 16 illustrates the predicted energy consumptions at both baseline and reporting periods based on the established chaining model with a 95% confidence interval.

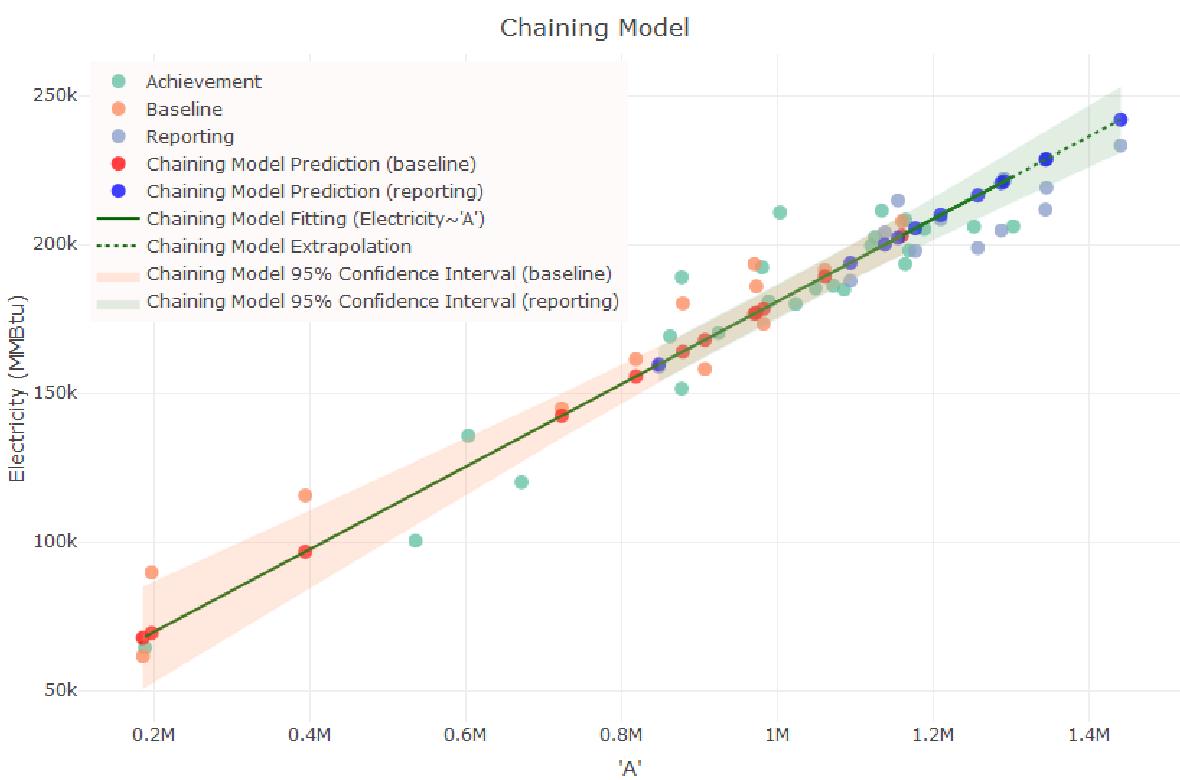

Figure 16 Baseline and Reporting Periods Energy Consumption Prediction of Chaining Model

With the predicted values for the baseline period and reporting period, a comparison of the predicted values with the observed electricity consumption can be made. Here we compare not only the chaining model predictions with the observations made during the baseline period, but also the chaining model predictions with the observations made during the reporting period with the "A" observed during the two periods of time accordingly. Two t-tests were conducted, and a p-value of 0.74 is obtained for baseline period observations and of 0.51 for reporting period observations, which indicate that the predicted energy consumptions and the observed ones are not statistically significantly different for both baseline period and reporting period. There is therefore no evidence that the facility's energy consumption performance is statistically significantly different between the baseline period and the achievement period, or between the achievement period and the reporting period.

5. Discussion

5.6. Model Comparison and Model Selection

5.6.1. Comparison of Energy Performance Improvement Percentage

Based on the four adjustment methods proposed by the SEP M&V Protocol, the SEP Energy Performance Indicator (SEnPI) and the Energy Performance Improvement (EPI) percentage can be calculated. Table 4 provides an overview of the four approaches used for the energy performance improvement evaluation as well as the obtained results which turned out to be very different among the four models.

Table 4 Energy Performance Indicator and Energy Performance Improvement Percentage Estimates Obtained by the Four Adjustment Model Methods

Method	Regression model built with data from	Coefficient of "A": $\beta_{"A"}$	Intercept: β_0	R^2	t-test p-value	SEnPI	EPI (%)	ESP_{TD} (MMBtu)
Forecast	Baseline period	0.13	53567.33	0.95	0.28	0.96	4.14%	106,412.7
Backcast	Reporting period	0.11	6917.97	0.82	0.99	1.00	-0.06%	-1091.4
Standard Conditions	Baseline period + Reporting period	forecast + backcast	forecast + backcast	-	0.66	0.98	2.19%	47,892.1
Chaining	Achievement period	0.14	4211.16	0.89	0.74, 0.51	0.93	6.58%	142,248.0

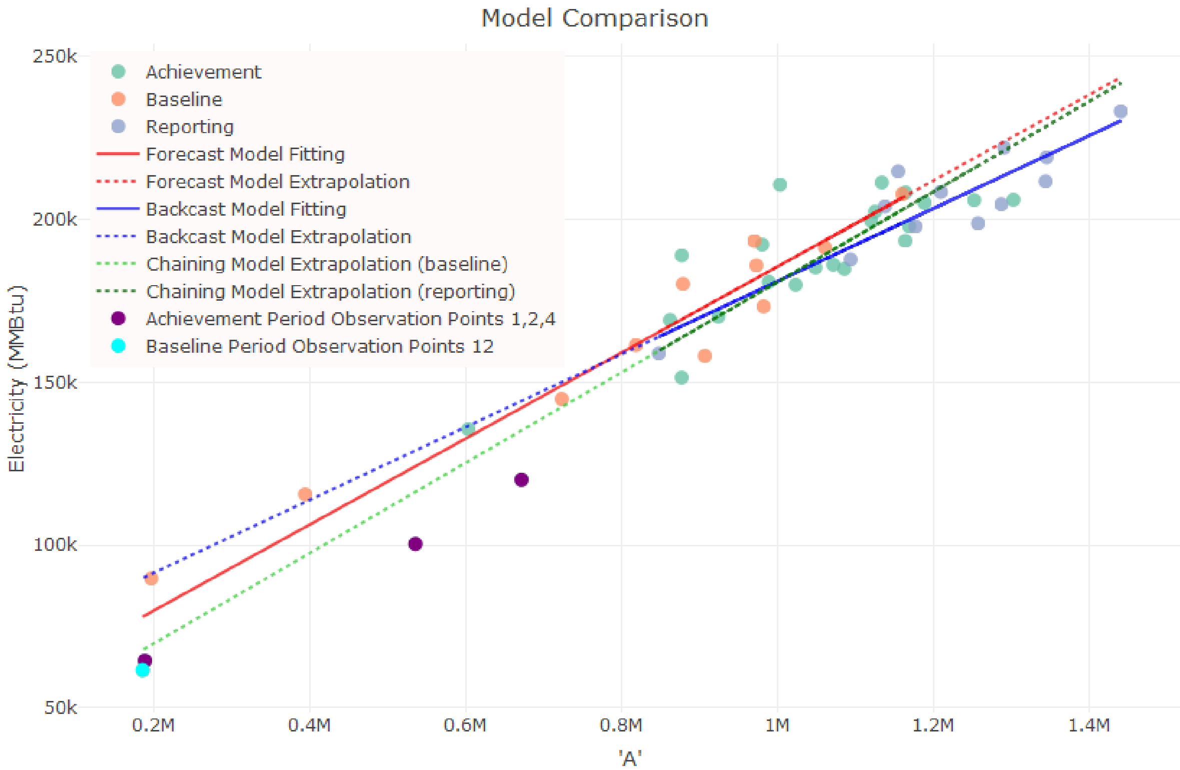

Figure 17 Adjustment Model Prediction Results Comparison

The evaluation results in Table 4 vary drastically according to the method employed. The primary issue that led to the differences in results is that the facility's production patterns changed from the baseline period to the reporting period (indicated by "A"). This is directly linked to the important change in the facility's energy consumption behavior between the two periods of time.

In fact, the baseline period production level is relatively low (low value range of "A") during the 12 months and contains no information indicating the facility's energy consumption behavior with high production level. Therefore, the forecast model built based on baseline period data is not able to predict accurately the facility's energy consumption during the reporting period, which corresponds to higher production level (see important uncertainty around the predicted reporting period consumption in Figure 10).

In similar ways, when building the backcast model, all the learning data are based on high production pattern (high values of "A" observed during the reporting period) and cover a narrow range of values for the relevant variable "A", which makes the model hard to predict the consumption behavior at baseline period with low production pattern (see large confidence interval in Figure 13). In the contrary, the intermediate period has two years of observations (24 points) which encompasses a large range of values of the relevant variable "A". In this case, it is more appropriate to adopt the standard condition model to perform the energy performance evaluation, which employs the fitting and the extrapolation of both the forecast model and the backcast model (see Figure 15) to obtain estimated energy consumption during the achievement period.

The chaining model is the only model that takes other data than those from the baseline and reporting periods into account in the performance evaluation. Only the achievement period's observations were used to establish the regression model. There are several points of low production (achievement period points 1, 2, 4 in Figure 17) that have a large influence on the fitting of the chaining model's slope and intercept (red line), which make the model yield more savings than the forecast and backcast models when comparing the baseline electricity consumption to the reporting period electricity consumption. The supplemental data provided by the intermediate period can represent neither the baseline period performance nor the reporting period performance. The difference in the intercept estimates (β_0) leads to biased comparison result which makes the chaining model score the best improvement evaluation (EPI) among all the methods considered in this study case. This approach is thus more appropriate to use when some changes irrelevant to the facility's energy performance occur (such as moving from one site to another, which is known as "static factors") during the achievement period that have a substantial impact on the energy consumption. In this case, the baseline observations and reporting period observations are not comparable and difficult to adjust (with non-routine adjustments). For example, the fact of moving from one site to another may help to reduce energy consumption, however the effect is confounded with the other actions taken to improve the energy performance and hard to be isolated. This is why an intermediate model is sometimes required to complete the evaluation.

As the average value of relevant variable "A" is higher during the reporting period, and the energy consumption has positive correlation with the value of "A", the forecast model estimating the energy consumptions at high level of "A" led to higher energy savings estimation than that obtained by the backcast model estimating the energy consumptions at low level of "A".

The most appropriate method should not be chosen based on the evaluation results, but on how the facility evaluates its future relevant variable level. If the future relevant variable level (range of "A") is closer to the reporting period, then the forecast method makes more sense. Likely, if the future level of "A" shares more similarity to the baseline period, then the backcast method is recommended. In the case that the facility has no clear idea of how the future level of "A" would be, the standard condition model would appear to be the most appropriate method, for it covers both low and high values of "A" and using no achievement period's consumption data to evaluate the energy performance improvement which could be confusing sometimes.

For facilities encountered similar situation as the study case, as the energy consumption range is changed from baseline period to reporting period due to the variation of the relevant variable(s), the adjustment method should be chosen with precautions since it can easily introduce bias in the energy performance evaluation which may lead to inappropriate conclusions.

Table 5 summarizes the interpretation of the energy performance evaluation results obtained by the four adjustment models. Guidance on how to choose the most appropriate adjustment model according to the available data and measurement conditions is also provided.

Table 5 Adjustment Models and Recommended Conditions to Use for Energy Performance Evaluation

Adjustment Model	Interpretation	Recommended Conditions to Apply
Forecast	Energy saved by the improvement actions during the reporting period (reporting period savings).	• If baseline period energy performance model can be extrapolated. • If reporting period measuring conditions (e.g. production level, weather conditions) are desired as the evaluation basis. • Usually used as the default method.
Backcast	Energy that could have been saved with the improvement actions during the baseline period (lost saving opportunity).	• If inappropriate operation schedule or habits were observed in baseline period that logical correlation between relevant variables and energy performance indicator cannot be established. • If reporting period energy performance model can be extrapolated. • If baseline period measuring conditions (e.g. production level) are desired as the evaluation basis.
Standard Conditions	Energy that could be saved by the improvement actions if the production activity is under the standard conditions.	• If important changes occur in relevant variable(s). • If some specific measuring conditions are considered more appropriate for the energy performance evaluation such as closer to the future operating conditions.
Chaining	Energy that could be saved by the improvement actions if no energy performance-unrelated changes have occurred that influence the facility's energy consumption.	• If reporting period and baseline period measuring conditions are changed (such as site change) that cannot be filtered by the adjustment model. • If the consumption behaviors during the two evaluation periods are not comparable due to the important difference between the relevant variable range (low value range vs high value range, with difference larger than 3 times of the standard deviation). • If non-representative consumption data were collected (a few identified outliers) that affect the energy performance evaluation. • If intermediate period data cover both low and high value range of relevant variables, chaining allows broader applications. Performance during

		intermediate period will affect evaluation results. • Generally, performance of later years (closer to reporting period) reflects better the results of saving actions.

5.7. New Energy Savings Indicator

Even if the relevant variable's influence on energy performance improvement could be isolated completely through modeling, the values on which a standard conditions model should be built for the performance evaluation depend on the facility's predictable future or average relevant variable values. If the energy consumption is positively correlated to the relevant variable, the use of high relevant variable value range would result in higher energy savings and a smaller SEnPI. To avoid the confusion and the bias introduced by the adjustment model, we propose an alternative indicator that could adjust the evaluation of energy performance and is independent of the relevant variable level, such as:

Equation 5
$$\widehat{\mathbf{ESP}} = \frac{1}{2}\left(\mathbf{ECP}(\boldsymbol{\Sigma})_{\mathbf{r}}^{\mathbf{o}} + \mathbf{ECP}(\boldsymbol{\Sigma})_{\mathbf{r}|\mathbf{b}}^{\mathbf{a}} - \mathbf{ECP}(\boldsymbol{\Sigma})_{\mathbf{b}|\mathbf{r}}^{\mathbf{a}} - \mathbf{ECP}(\boldsymbol{\Sigma})_{\mathbf{b}}^{\mathbf{o}}\right)$$

Table 6 Top-Down Energy Saving Estimates Comparison

Method	$\mathrm{ESP}_{\mathrm{TD}}$ (MMBtu)
Forecast	106,412.7
Backcast	-1091.4
Standard Conditions	47,892.1
Chaining	142,248.0
Proposed $\widehat{\mathrm{ESP}}$	52,660.6

Table 6 lists the energy saving estimates resulted from all the four adjustment methods as well as the suggested estimator. The proposed estimator $\widehat{\mathrm{ESP}}$ provides closer estimate to the standard conditions results. In the case that the standard conditions are hard to define, this estimator could provide an estimated saving value that appears to be more representative than the ones provided by forecast or backcast models alone.

5.8. Energy Performance Indicator with Confidence Interval

Applying a confidence interval to the energy savings value is usually helpful to understand the possible range of the energy performance improvement value obtainable based on the limited available data. The 95% confidence interval can be interpreted as that for a given SEnPI value, 95% of the time the true mean SEnPI will be between the lower and upper values seen in Table 7, which provides the estimated 95% confidence range for each energy performance improvement estimate resulting from the application of the four methods.

Table 7 shows that most of the confidence intervals are relatively large which coincides with the conclusion of the z-tests performed for each model: the energy performance in the baseline period and in the reporting period is not statistically significantly different.

Table 7 Energy Performance Indicator, Energy Performance Improvement Percentage and Top-Down Energy Saving Estimates Obtained by the Four Adjustment Models with 95% Confidence Interval

Method	SEnPI	EPI (%)	ESP_{TD} (MMBtu)
Forecast	0.959 [0.909; 1.014]	4.144 [-1.436; 9.143]	106,412.7 [-348,55.1; 247,680.6]
Backcast	1.001 [0.885; 1.116]	-0.059 [-11.649; 11.531]	-1091.4 [-217,039.0; 214,856.1]
Standard Conditions	0.978 [0.869; 1.099]	2.193 [-9.905; 13.103]	47,892.1 [-205,183.1; 300,967.3]
Chaining	0.934 [0.851; 1.023]	6.584 [-2.322; 14.863]	142,248.0 [-48,963.0; 333,459.0]

6. Conclusion

The Superior Energy Performance (SEP) Measurement and Verification (M&V) Protocol provides requirements for determining facility-wide energy performance improvement in support of the ISO 50001 based SEP program. The SEP M&V Protocol instructs users to adjust energy consumption data in either a baseline, reporting, or intermediate period to account for variations in relevant variables, which affect energy consumption at an organization but they themselves are independent of implemented energy performance improvement actions. The SEP M&V Protocol specifies four energy consumption adjustment modeling methods for use; forecast, backcast, standard conditions, and chaining.

Application of a single set of energy consumption and relevant variable data from a manufacturing facility to the four different energy consumption adjustment modeling methods produced four different energy savings estimates that range from -1091.4 to 142,248.0 MMBtu, and four SEP Energy Performance Indicator (SEnPI) estimates ranging from 0.93 to 1.00, which led to different energy performance improvement conclusions. Variation in the energy savings estimates is the result of inevitable changes in operation and conditions between the baseline and reporting periods, which affect the evaluation results significantly. The lack of agreement in the obtained energy savings values, while all meeting the requirements of the SEP M&V Protocol, indicates that additional context and analysis is needed to understand which modeling method, and subsequent result, best represents the actual energy performance improvement of an organization.

A variety of statistical tests are applied to the data and energy savings results to reveal which of the four methods best describes and reflects the facility’s energy consumption evolution and whether the obtained energy performance improvement percentages make sense based on the facility’s operation patterns.

Ultimately, the selected method should include a modeling period that encompasses relevant variable values that are close to expected future levels, so that the evaluation results are more meaningful and indicative of the facility's future energy savings. In the absence of such knowledge of future relevant variable levels while important variation of the relevant variable is observed during the baseline and reporting periods, the standard conditions approach is well adapted to determine meaningful energy savings values as the method may be set up with a vast range of relevant variable values reflected in both the baseline and reporting periods.

Lastly, an alternative average energy saving estimator is proposed which allows to adjust the evaluation of energy performance and is independent of the relevant variable level. The interpretation of 95% confidence interval reveals more context to what calculated energy savings values mean and the level of reliability with which the estimated mean value should be interpreted.

Future work will include expansion of this analysis to include data from additional facilities to validate that the challenges presented in this report are universal. With additional analysis, a methodology by which to select the appropriate modeling method after original data have been collected will be developed and proposed.

7. Bibliography

US. Department of Energy, 2017, Superior Energy Performance Measurement & Verification Protocol, Washington DC

International Standards Organization (2016) ISO 17747 – Determination of energy savings in organizations

International Standards Organization (2014) ISO 50015 – Measurement and verification of organizational energy performance – General principles and guidance

International Standards Organization (ISO) (2011) ISO 50001 – Energy management system standard – Requirement with guidance for use

Goldstein, D. and Almaguer, J. (2013) Developing a Suite of Energy Performance Indicators (EnPIs) to Optimize Outcomes, 2013 ACEEE Summer Study on Energy Efficiency in Industry, Niagara Falls, NY

Goldstein, D., McKane, A., and Desai, D. (2011) ISO 50001: Energy Management Systems: A Driver for Continual Improvement of Energy Performance, 2011 Summer Study on Energy Efficiency in Industry, Washington D.C.

8. Nomenclature

E(*)	Quantity of energy of an unspecified type
ECP(*)	Primary energy consumption of an unspecified energy type
ECP(Σ)	Primary energy consumption of all energy types
$\mathrm{ECP}(\Sigma)_{\mathrm{b}}^{\mathrm{o}}$	Observed (actual) baseline period energy consumption
$\mathrm{ECP}(\Sigma)_{\mathrm{r}}^{\mathrm{o}}$	Observed (actual) reporting period energy consumption of all energy types
$\mathrm{ECP}(\Sigma)_{\mathrm{b}\vert\mathrm{r}}^{\mathrm{a}}$	Modeled baseline period primary energy consumption adjusted to reporting period conditions
$\mathrm{ECP}(\Sigma)_{\mathrm{r}\vert\mathrm{b}}^{\mathrm{a}}$	Modeled reporting period primary energy consumption adjusted to baseline period conditions
$\mathrm{ECP}(\Sigma)_{\mathrm{b}\vert\mathrm{s}}^{\mathrm{a}}$	Modeled baseline period primary energy consumption adjusted to standard conditions
$\mathrm{ECP}(\Sigma)_{\mathrm{r}\vert\mathrm{s}}^{\mathrm{a}}$	Modeled reporting period primary energy consumption adjusted to standard conditions
ECP(Σ)i	Intermediate period total primary energy consumption
$\mathrm{ECP}(\Sigma)_{\mathrm{b}\vert\mathrm{i}}^{\mathrm{a}}$	Modeled baseline period primary energy consumption adjusted to intermediate period conditions
$\mathrm{ECP}(\Sigma)_{\mathrm{i}\vert\mathrm{r}}^{\mathrm{a}}$	Modeled intermediate period primary energy consumption adjusted to reporting period conditions
ESP(*)	Primary energy savings of an unspecified energy type